Molecular target identification of two *Coleus amboinicus* leaf isolates toward lung cancer using a bioinformatic approach and molecular docking-based assessment

Winarto Haryadi\(^1\)*, Kasta Gurning\(^1,2\), Endang Astuti\(^1\)

\(^1\)Departemen of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Yogyakarta, Indonesia. 
\(^2\)Departemen of Pharmacy, Sekolah Tinggi Ilmu Kesehatan Senior Medan, Medan, Indonesia.

**ABSTRACT**

Lung cancer raises a serious concern because of its position as the second global deadliest disease in both men and women. Conventional treatments emphasize chemotherapy and radiation or their combination, which is considered to be ineffective due to harmful side effects in patients. In this work, we did a bioinformatic investigation using the pharmacological network and molecular docking approaches to evaluate the potency of natural compounds isolated from the leaves of *Coleus amboinicus*, Lour, i.e., 16-hydroxy-7α-acetoxyroyleanone and 16-acetoxy-7α-hydroxyroyleanone. The activity of both active compounds against lung cancer was predicted using disease databases and genes. A total of 77 core targets were identified from the analysis using STRING and built using Cytoscape. Gene ontology, and kyoto encyclopedia of genes, and genome analysis of cancer pathways targeting PGTS2 and peroxisome proliferator-activated receptor gamma were employed because of their critical functions in cancer therapy. The validated molecular docking analysis illustrates the possibility of interactions that occur between active compounds of target proteins in the treatment of lung cancer. This research has not yet been able to demonstrate its potential in the treatment of lung cancer, and further research is needed to prove it through *in vitro* and *in vivo* examinations before pre-clinical and clinical tests in the future.

**INTRODUCTION**

Lung cancer is the deadliest cancer because it is one of the most chronic types of cancer in the world [1–3] and can occur in men and women [4,5]. This disease based on basal epithelial cells is classified into two types, namely non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC) [6]. NSCLC accounts for around 85% of lung cancer cases, with details of 40% adenocarcinoma, 5%–10% large cell carcinoma, and the remaining 30% originating from squamous cells. Meanwhile, SCLC will account for 15% of all cancers consisting of normal, undifferentiated small cells by 2030 [3,7].

Lung cancer could be caused by tobacco smoking, environmental exposures (such as second-hand smoke, air pollution, and radon), exposure to asbestos and arsenic, and epigenetic changes [8]. Commonly used treatments for lung cancer include surgery, radiation therapy, chemotherapy, molecular targeted therapy including epidermal growth factor receptors or anaplastic lymphoma kinase inhibition, and immunotherapy [9,10]. Although advanced treatment methods have been developed, the prognosis for lung cancer remains very unsatisfactory. Recent research reports that chemical compounds extracted from natural products are potential and effective for the treatment of lung cancer. Drug formulations from natural ingredients, namely paclitaxel, doxorubicin, and camptothecin, have been reported to have high efficacy in the treatment of lung cancer [11]. Therefore, targeted cancer therapy with specific signaling using natural compounds is promising and is being developed into pre-clinical and clinical trials [12]. Specifically, treatment with various combinations of natural compounds shows good effectiveness and safety levels in the treatment of lung cancer, which were tested *in vivo* [13], so they are promising in the development of bioactive compounds derived from natural ingredients.
The royleanone compounds and their derivatives from natural materials have been successfully isolated, including 7α-acetoxyroyleanone and Horminin from the *Peltodon longipes* plant, and showed a cytotoxic effect on pancreatic cancer cells (MIA PaCa-2) and melanoma cancer cells (MV-3) [14,15]. Hormone compounds from the plant *Salvia lachnocalyx* show cytotoxicity against breast cancer cells (MCF-7) and human chronic myelogenous leukemia (K-562) [16]. The compounds 16-hydroxy-7α-acetoxyroyleanone and 16-acetoxy-7α-hydroxyroyleanone were successfully isolated from the leaves of the *Coleus amboinicus*, Lour., plant and showed potential as antioxidants [17,18]. The chemical structure of 16-hydroxy-7α-acetoxyroyleanone and 16-acetoxy-7α-hydroxyroyleanone is shown in Figure 1. These compounds belong to royleanone derivatives, which are reported to have potential in cancer treatment. Based on the previous information, we will continue to study the potential application of both royleanone derivatives in the treatment of lung cancer using bioinformatic approaches and network pharmacology as initial assays before further developing *in vitro* and *in vivo* tests.

**MATERIALS AND METHODS**

**Materials**

In general, the tools used in molecular docking include computers (Intel Core i5-10400F, 16 GB RAM DDR4, 256GB SSD NVME, and VGA AMD Radeon RX6600) with a Windows 10 Home operating system, Auto Dock Tools 1.5.6, PyRx 0.8, PyMOL 2.3, GaussView 5.0 (MDL Information Systems, Inc.), and Discovery Studio 21.0 Client (DSV 19.0) software. The natural compounds that were studied in this study through a bioinformatic approach were 16-hydroxy-7α-acetoxyroyleanone and 16-acetoxy-7α-hydroxyroyleanone. The bioinformatics approach to the isolated compounds used the Swiss target prediction database, Cytoscape 3.10.1, Online Mendelian Inheritance in Man (OMIM), DisGeneNet, therapeutic targets database (TTD), and GeneCards.

**Isolated active compound**

Two active compounds, named 16-hydroxy-7α-acetoxyroyleanone and 16-acetoxy-7α-hydroxyroyleanone were isolated from the ethyl acetate extract of the leaves of the *C. amboinicus*, Lour., plant. The ethyl acetate extract was obtained from the multilevel partitioning of the methanol extract with n-hexane, chloroform, and ethyl acetate, as reported earlier [17,18]. Structures of isolated compounds were built with GaussView 5.0 (MDL Information Systems, Inc.) and optimized with Gaussian 09W using the density function theory (DFT) method with the B3LYP hybrid function and 3-21G(d,p) basis set [19].

**Identification of the potential activity of active compounds against lung cancer**

Active compounds resulting from 3-D optimization were opened with ChemDraw to obtain SMILE data for predicting target genes and molecular docking. SMILE active compound structure data was uploaded to the Swiss target prediction database (http://www.swisstargetprediction.ch/).

The predicted target gene data was downloaded in CSV format, and the data was filtered and integrated using Microsoft Excel software. Predicted target components were imported to UniProt for the normalization process, followed by the restriction to the human species. Afterward, all target proteins were retrieved and corrected according to common names. Anticancer lung targets of the active compounds were imported into Cytoscape (3.10.1) to generate “target-active compounds” tissue. The link between the active compound and the target is called a node, and the correlation between the active compound and the target is called an edge. Information on proteins involved in lung cancer was obtained using the National Center for Biotechnology Information GeneCards database (https://www.genecards.org/), TTD (https://db.idrblab.net/ttd/), OMIM (https://www.omim.org/), and DisGeNET using the keyword “lung cancer.” Both information on active compound components and treatment target components in lung cancer were processed using VENNY 2.1 (https://bioinfogp.cnb.csic.es/tools/venny/index.html) to obtain Venn diagrams of active compounds and with lung cancer as the component targets [12,20,21].

**Protein-protein interaction (PPI) analysis and screening**

PPI determines the interaction relationship between two or more proteins based on biochemistry, hydrophobicity, and electrostatics. Protein is an important part of biological processes that occur in living things, both in healthy and diseased conditions. Gene and protein interconnections are represented by PPI, as this network is an important part of bioinformatics research [20]. The protein that intersects the active compound and lung cancer is uploaded to the STRING online site (https://string-db.org/) with the settings *Homo sapiens* with a confidence level of 0.4, and other parameters were selected as default. PPI relationships with node1, node2, and the combined score from the export file are imported into Cytoscape 3.10.1 to form an interaction network. The results of the analysis of the top target were carried out using the CytoHubba plugin to obtain the highest protein with the highest 20-degree value for the next stage [12,21].

**Gene Ontology (GO) and Kyoto Encyclopedia of Genes, and Genomes (KEGGs) pathway analysis**

GO and KEGG analyses were performed with the approach of biological processes, cellular components, gene
functionality, cellular components, and molecular functions. The results of bioinformatic analysis of STRING data [12,21] were obtained by importing active compounds intercept targets with lung cancer to result in the molecular mechanisms of active compounds in the treatment of lung cancer.

Molecular docking approach

Molecular docking is a rational drug design method based on exploring the interaction between ligands (active compounds) and receptors (proteins). The interactions were studied to predict the binding affinity of active compounds for proteins, as well as the intermolecular patterns conferring biological activity [22,23]. Molecular docking could be carried out using AutoDock Vina, developed by the Scripps Research Institute, for semi-flexible molecular docking computing. AutoDock Vina uses complex gradient algorithms and multi-threaded techniques to produce predictions that are more accurate and faster when compared to AutoDock 4 [21]. The optimized active compounds were tethered to three cancer proteins obtained from the KEGG cancer pathway. The protein codes were traced to the PDB database (https://www.rcsb.org/) and selected specifically for lung cancer. The best protein crystal structures were selected, and the PDB database was downloaded. Protein PDB files of active compounds and ligand molecules are imported into the AutoDock tool for molecular docking. AutoDock Vina was run to bind the active compound treated to the target protein ten times, and the lowest binding energy for each docking was taken as the final result. Complexes were then observed and plotted using PyMOL [21].

RESULTS AND DISCUSSION

Optimization of active compounds

Two active compounds, namely 16-hydroxyroyleanone-7α-acetoxyroyleanone and 16-acetoxy-7α-hydroxyroyleanone, were prepared by optimizing their structure to obtain the most stable energy with Gaussian 09W using the DFT method with B3LYP hybrid function and 6-31G(d,p) basis set. Their structures are then validated with AutoDockTools, as shown in Table 1.

The optimizing process of the compounds was carried out to obtain thermal stability and molecular charge as well as their biochemical behavior based on quantum mechanics calculations [24]. The results of molecular optimization are expected to have the same molecular structure in the human body.

Identification of active compounds against lung cancer

The Swiss Target Prediction results had the densest 50 targets from 16-acetoxy-7α-hydroxyroyleanone (Compound 1) and 54 targets from 16-hydroxy-7α-acetoxyroyleanone (Compound 2). At the same time, the combined prediction of both active compounds produced 104 targets. Predictions of 104 total targets were analyzed by eliminating targets so they were not doubled to obtain a total of 77 targets. The resulting active compound-target network was formed using Cytoscape 3.10.0 software (Fig. 2a). Based on validation results from various databases, 18,404 protein targets associated with lung cancer were identified. Both 16-acetoxy-7α-hydroxyroyleanone and 16-hydroxyroyleanone-7α-acetoxyroyleanone have intercepted 77 potential targets that are predicted to be related to lung cancer proteins (Fig. 2b).

Target analysis of PPI networks

The total targets predicted (77 targets) that had intercepts with lung cancer were analyzed and then imported into a string to build a PPI network by selecting *Homo sapiens* organisms with a confidence of 0.40 (Fig. 3a). Intercepts of active target-compounds were imported into Cytoscape 3.10.0 to construct and obtain a network map of potential target interactions (Fig. 3b). The results of the analysis showed that there were 77 nodes (proteins) and 214 edges (interactions). Based on the linkages informing the PPI, the 10 highest (highest to lowest) interaction targets include prostaglandin-endoperoxide synthase-2 (PTGS2), peroxisome proliferator-activated receptor gamma (PPARG), ACE, MMP2, ITGB1, SERPINE1, REN, ITGB3, PPARA, KDR and they were considered to have the best relationship prediction for lung cancer.

<table>
<thead>
<tr>
<th>Active compounds</th>
<th>Energy (kJ/mol)</th>
<th>Dipole moment</th>
<th>Compounds name</th>
</tr>
</thead>
<tbody>
<tr>
<td>16-acetoxy-7α-hydroxyroyleanone (Syn. 16-hydroxy-7-O-acetylhorminone)</td>
<td>−1,300.609</td>
<td>3.4830</td>
<td>C_{22}H_{30}O_{6}</td>
</tr>
<tr>
<td>16-hydroxy-7α-acetoxyroyleanone (Syn. 16-acetoxyhorminone)</td>
<td>−1,301.125</td>
<td>2.8532</td>
<td>C_{22}H_{30}O_{6}</td>
</tr>
</tbody>
</table>

Table 1. The active compounds isolated from the ethyl acetate extract of *C. amboinicus*, Lour leaves.
GO and the KEGG pathway analysis

GO and KEGG pathways were analyzed to explore the mechanisms of 77 targets in the treatment of lung cancer using a bioinformatics database (ShinyGO 0.77). GO and KEGG results of active compounds against lung cancer are displayed in Figure 4. The analysis is carried out on biological processes, cellular components, molecular functions, and the KEGG pathway of the potential targets of active compounds in the treatment of lung cancer, which are shown in Figure 4a–d, respectively.

The enrichment results identified the 20 pathways associated with active compound targets in the treatment of lung cancer that have been analyzed according to the KEG pathways from potential targets of active compounds. The targets associated with each of these pathways are described in Table 2. Based on the cancer pathway, there are 14 targets related to the cancer pathway, i.e., MMP2, PTGER2, ITGAV, PPAR, PTGER4, GSTM1, MMP1, NOS2, PTGS2, PIM1, EDNRB, ITGB1, BCL2, and AGTR1. The two targets with the highest interaction between active compounds and lung cancer in the cancer pathway are found in PTGS2 and PPAR.

In silico and molecular docking

The two best target interactions of active compounds for the treatment of lung cancer are selected based on the results of KEGG pathways analysis, namely PTGS2 and

Figure 2. Identification of active-target compounds: (a) Active-target combined network with Cytoscape 3.10.0 and (b) intercept Venn diagram of each compound against lung cancer targets.

Figure 3. Analysis of PPI: (a) PPI network built from intercept prediction of active-target compounds using strings, and (b) proteins (nodes) that are linked by interactions (edges) with one another, and blue color represents the highest PPI top 10.
Figure 4. GO and KEG analysis of active compounds for the treatment of lung cancer was built using the ShinyGO 0.77 database: (a) Biological processes, (b) cellular components, (c) molecular functions, and (d) KEGG pathways from potential targets of active compounds in the treatment of lung cancer.

Table 2. Pathways related to active compounds and lung cancer.

<table>
<thead>
<tr>
<th>Entry</th>
<th>Pathways</th>
<th>Term</th>
<th>Targets</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Renin-angiotensin system</td>
<td>hsa04614</td>
<td>REN, ACE, AGTR2, CTSA, MME, AGTR1</td>
</tr>
<tr>
<td>2.</td>
<td>Circadian rhythm</td>
<td>hsa04710</td>
<td>PRKAB1, PRKAG1, PRKAA1, PRKAA2</td>
</tr>
<tr>
<td>3.</td>
<td>Hypertrophic cardiomyopathy</td>
<td>hsa05410</td>
<td>ITGB4, PRKAB1, ITGAV, ITGB7, ACE, PRKAG1, PRKAA1, PRKAA2, ITGB1, ITGA4, ITGB3</td>
</tr>
<tr>
<td>4.</td>
<td>PPAR signaling pathway</td>
<td>hsa03320</td>
<td>FABP4, PPARG, FABP5, MMP1, FABP3, PPARA</td>
</tr>
<tr>
<td>5.</td>
<td>Arrhythmogenic right ventricular cardiomyopathy</td>
<td>hsa04542</td>
<td>ITG4, ITGAV, ITGB7, ITGB1, ITGA4, ITGB3</td>
</tr>
<tr>
<td>6.</td>
<td>Glucagon signaling pathway</td>
<td>hsa04922</td>
<td>PYGL, PRKAB1, PRKAG1, PRKAA1, PRKAA2, PPARA</td>
</tr>
<tr>
<td>7.</td>
<td>Adipocytokine signaling pathway</td>
<td>hsa04920</td>
<td>PRKAB1, PRKAG1, PRKAA1, PRKAA2, PPARA</td>
</tr>
<tr>
<td>8.</td>
<td>Renin secretion</td>
<td>hsa04924</td>
<td>PTGER2, REN, ACE, PTGER4, AGTR1</td>
</tr>
<tr>
<td>9.</td>
<td>ECM-receptor interaction</td>
<td>hsa04512</td>
<td>ITG4, ITGAV, ITGB7, ITGB1, ITGA4, ITGB3</td>
</tr>
<tr>
<td>10.</td>
<td>Insulin resistance</td>
<td>hsa04931</td>
<td>PYGL, PRKAB1, PRKAG1, PRKAA1, PRKAA2, PTPN1, PPARA</td>
</tr>
<tr>
<td>11.</td>
<td>Dilated cardiomyopathy</td>
<td>hsa04541</td>
<td>ITG4, ITGAV, ITGB7, ITGB1, ITGA4, ITGB3</td>
</tr>
<tr>
<td>12.</td>
<td>Fluid shear stress and atherosclerosis</td>
<td>hsa04518</td>
<td>MMP2, ITGAV, KDR, GSTM1, PRKAA1, PRKAA2, BCL2, ITGB3</td>
</tr>
<tr>
<td>13.</td>
<td>Longevity regulating pathway</td>
<td>hsa04211</td>
<td>PRKAB1, PPAR, PRKAG1, PRKAA1, PRKAA2</td>
</tr>
<tr>
<td>14.</td>
<td>Neuroactive ligand-receptor interaction</td>
<td>hsa04080</td>
<td>OPRD1, PTGER2, THRA, AVPR1A, PTGER4, OXTR, CCKBR, C5AR1, AVPR2, GRIN1, AGTR2, EDNRB, THRB, OPRM1, GPR35, AGTR1</td>
</tr>
<tr>
<td>15.</td>
<td>Apelin signaling pathway</td>
<td>hsa04371</td>
<td>SERPINE1, PRKAB1, PRKAG1, NOS2, PRKAA1, PRKAA2, AGTR1</td>
</tr>
</tbody>
</table>

Continued
The PTGS2 plays an important pathological role in the treatment of chemotherapy-resistant cancer. Moreover, PTGS2 has a role in increasing the response of bodies that are resistant to chemotherapy treatment [25, 26]. Therapy for non-drug-resistant cancer cells could be done by inhibiting the expression of PTGS2. Inhibition of PTGS2 expression will suppress the proliferation, migration, and invasion of cancer cells, as well as modulate the immune response by impairing cell differentiation and suppressing the occurrence of metastases [27]. On the other hand, the PPARG target is the most abundant subtype expressed in adipose tissue with two isoforms. The inhibition of the PPARG target is able to inhibit cell proliferation, induce cell cycle arrest, apoptosis of multiple cancer cells, increase adhesion between cells, and immobilize the tumor microcell environment, which promotes both transcription and protein levels and prevents metastases [28,29]. In addition, PPARG also plays an important role in regulating the expression of various genes regarding glucose and lipid metabolic homeostasis, adipogenesis, and inflammation [30]. These two targets of active compounds for the treatment of lung cancer have two opposite directions: PPARG inhibits apoptosis, signaling, proliferation, and metastases, thus overcoming the occurrence of resistance and increasing the activity of cancer treatment [31,32].

Besides, suppressing or inhibiting the expression of the PTGS2 target resulted in a maximum treatment and inhibition of cancer invasion and metastases [31,33]. In silico testing of the active compounds was carried out on target proteins that have been reported as a therapy in cancer treatment. The protein codes used in this approach are 6GES (PTGS2 protein) and 5ZMA (PPARG protein) (Table 3 and Fig. 5).

**Table 3.** Strength of interaction between active compounds and potential lung cancer target proteins.

<table>
<thead>
<tr>
<th>Entry</th>
<th>Pathways</th>
<th>Term</th>
<th>Targets</th>
</tr>
</thead>
<tbody>
<tr>
<td>16.</td>
<td>Insulin signaling pathway</td>
<td>hsa04910</td>
<td>PYGL, PRKAB1, PRKAG1, PRKAA1, PRKAA2, PTPN1</td>
</tr>
<tr>
<td>17.</td>
<td>Focal adhesion</td>
<td>hsa04510</td>
<td>ITGB4, ITGAV, KDR, ITGB7, ITGB1, ITGA4, BCL2, ITGB3</td>
</tr>
<tr>
<td>18.</td>
<td>Calcium signaling pathway</td>
<td>hsa04020</td>
<td>AVPR1A, OXTR, NOS2, CCKBR, GRIN1, EDNRB, AGTR1</td>
</tr>
<tr>
<td>19.</td>
<td>PI3K-Akt signaling pathway</td>
<td>hsa04151</td>
<td>ITGB4, ITGAV, KDR, ITGB7, PRKAA1, PRKAA2, ITGB1, ITGA4, BCL2, ITGB3</td>
</tr>
<tr>
<td>20.</td>
<td>Pathways in cancer</td>
<td>hsa05200</td>
<td>MMP2, PTGER2, ITGAV, PPARG, PTGER4, GSTM1, MMP1, NOS2, PTGS2, PIM1, EDNRB, ITGB1, BCL2, AGTR1</td>
</tr>
</tbody>
</table>

**CONCLUSION**

The reported bioinformatic study of two compounds isolated from the leaves of *C. amboinicus* Lour., namely 16-acetoxy-7α-hydroxyroyleanone and 16-hydroxy-7α-
acetoxyroyleanone, suggests potential application in the treatment of lung cancer. The results of integrated bioinformatic analysis of GO, KEGG, and cancer pathways have 77 targets, with PTGS2 and PPARG being the two main targets for treating lung disease. Validated molecular docking analysis revealed multiple beneficial interactions in the active sites of PTGS2 and PPARG receptors. These findings are important as initial information and need to be developed into further research stages. This research is still very limited because it was only carried out based on bioinformatic studies, so further testing needs to be done in vitro and in vivo regarding these two active compounds as therapeutic targets in the treatment of lung cancer.

ACKNOWLEDGMENT

The authors are grateful for the support of facilities from the Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Yogyakarta, while conducting this research.

AUTHOR CONTRIBUTIONS

All authors made substantial contributions to the conception and design, acquisition of data, or analysis and interpretation of data; took part in drafting the article or revising it critically for important intellectual content; agreed to submit to the current journal; gave final approval of the version to be published; and agree to be accountable for all aspects of the work. All the authors are eligible to be an author as per the International Committee of Medical Journal Editors (ICMJE) requirements/guidelines.

FINANCIAL SUPPORT

There is no funding to report.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

ETHICAL APPROVALS

This study does not involve experiments on animals or human subjects.

DATA AVAILABILITY

All data generated and analyzed are included in this research article.

PUBLISHER’S NOTE

This journal remains neutral with regard to jurisdictional claims in published institutional affiliation.

REFERENCES


How to cite this article: